图书情报工作
第 57 卷 第 20 期　2013 年 10 月

德国国家科技图书馆的发展思考与实践*

Jan Brase　Uwe Rosemann　Irina Sens(著)　顾立平(译)

［摘　要］简介科研数据应用的新发展以及科研数据权益管理的新挑战。描述德国国家科技图书馆的可扩展架构以及其中的科研数据入口网站 GetInfo、永久性识别服务 DOI、全球性数据引用服务 DataCite 等的实践成果,并讨论培养数据馆员的需求。这些服务可以增强对非文本资源的检索、标识数字数据资源、促使科研数据得到正式引用以及提升图书馆价值。

［关键词］数据引用　数字资源唯一标识符　科研数据集成入口　科研数据权益与管理　元数据框架
［分类号］G250.76
DOI:10.7536/j.issn.0252 -3116.2013.20.009

1　科技信息服务的挑战与未来图书馆界的发展

科技信息越来越庞大复杂,而且逐步趋向海量数据的集成。尽管知识传播仍然有赖于科学论文的形式,然而,若从整体的科研过程来看,它仅仅是科研数据管理流程的最后一道程序。

美国国家科学基金会在其报告《面向 21 世纪探索的网路基础设施》中提到:“科技研究和教育越来越向数据密集型的方向转变:以数字技术、仪器仪表和各种网络来收集、生成、共享和分析数据,并且进行扩散。全世界范围内的科学家和工程师们每天通过实验、观察和模拟,进行生产、获取、分析、整合和存储 T - bytes 级别的数字数据。由此而来,新的科学方法以观察和模拟生成数据的动态整合方式,智能地适应不断变化的条件,从而揭开新的知识”[1]。科研数据应用的崭新发展以及科研数据权益管理的全新挑战,引导着图书馆界迈向新的发展方向。

微软研究院的开放著作《第四范式:数据密集型科研探索》,特别注意到科学自身如何在 20 世纪末从实证主义转向数据密集。科学传播,包括同行评审,也在发生根本性的变化。由于费用、时效以及把实验数据和论文文献放在一起的需求等,数字图书馆的公共共享服务将会全面代替传统图书馆在收藏出版物上的角色”[2]。

科学信息现在不仅仅以一篇文章或者一本图书的形式来呈现。图书馆门户网站可以整合信息并且提供开发服务,从而持续作为用户获取信息的一个重要来源[3]。

图书馆用户最关心的是能否在查询信息后获得满意答复(无论他查找所获得的信息是论文、图书或者其他形式的内容,例如科研数据、三维模型、图片、视频等)。作为知识和信息服务的提供者,图书馆员的任务是提供用户服务,针对他们的提问给予合适的科学信息作为最为理想的答复。如果图书馆界想要持续履行这种信息供应的职责,就得释放不同形式内容的书目数据。所以,重点在于:

- 知识对象的显示依赖合适的元数据。这意味着:针对各个对象,通过各类信息,尽可能地丰富目前已有的元数据,以满足用户搜索。所以,在目录中与关键词相关的传统搜索方法需要扩展成为能够满足各种不同内容形式的新型搜索方法。
- 需要开发资源采购和建设的新方法,使之成为内容质量的保证。
- 参考和引用的支撑能力,是记载科学内容的出版物获得永久可用性的前提。为此,过去几年德国国家科技图书馆(Gemeinsamen Bibliotheksverbund,TIB)

* 原文以“Kommentar aus Bibliothekssicht”为题,发表在 *Zeitschrift für Bibliothekswesen und Bibliographie*,经过作者们授权,翻译并且略为调整后,以中文发布。本文系中国科学院国家科学图书馆指向性人才研究项目“科技信息政策研究与咨询”(项目编号:馆 1203)研究成果之一。
［作者简介］Jan Brase,德国国家科技图书馆 DOI 与 DataCite 负责人,博士,Email:jan.brase@tib.uni-hannover.de;Uwe Rosemann,德国国家科技图书馆与汉诺威大学图书馆馆长;Irina Sens,德国国家科技图书馆副馆长兼数字图书馆与知识服务部门主任,博士;顾立平,中国科学院国家科学图书馆副研究员,博士,E-mail:gulp@mail.las.ac.cn。
收稿日期:2013 - 09 - 14　修回日期:2013 - 10 - 08　本文起止页码:58 - 61　本文责任编辑:王传清

建立了 DOI 注册代理以及相关的国际协会(即 DataCite)。

可扩展性意味着图书馆能够直接处理若干海量信息以及那些传统上就已熟悉的内容载体,但是不必自己把持所有内容。图书馆一方面要考虑到存储海量大型科研数据集的技术要求;另一方面要考虑到专门存储和维护这些内容的认证机制和措施。

图书馆目录不应作为纯粹展示自己资源的窗口,而应该思考将之发展成为一种分布式延伸内容的门户。所谓"延伸内容",是指类似大数据集,或者被编目的视频和学习对象,提供与用户的研究相关的匹配结果。读者在顺利访问和获取内容后,通过稳定的链接,还可以延伸触及那些受到认证的专门存储有关内容的档案馆和数据银行。

上述这种可以持续深化发展的数字图书馆架构,如图 1 所示:

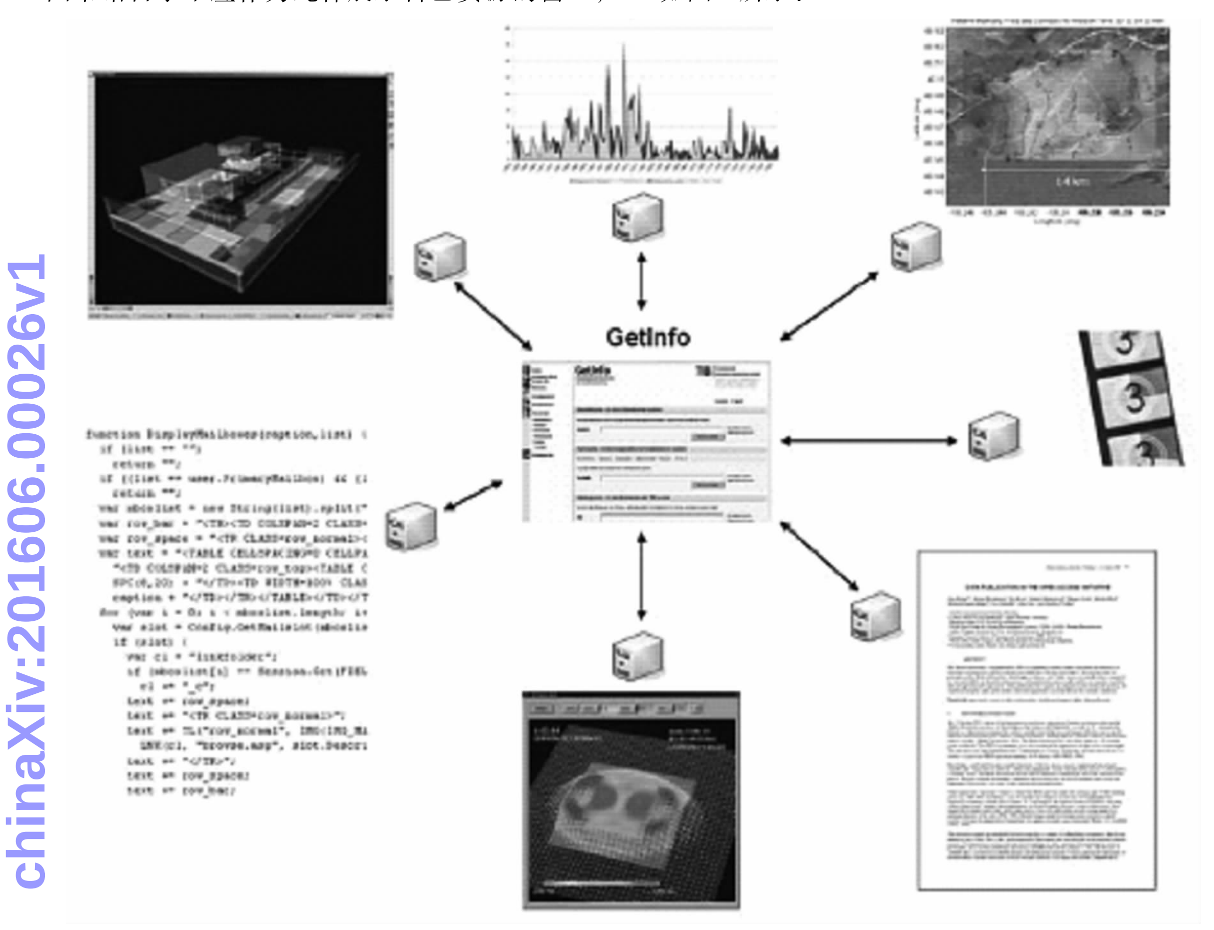


图 1 持续深化发展的图书馆入口网站(以 GetInfo 为例)

2 科研数据 GetInfo 入口网站:案例说明

德国国家科技图书馆是为建筑、化学、信息计量、数学和物理等专业学科而建立的国家级专业图书馆。它的 GetInfo 是专为自然科学与工程专业等科研信息而建立的入口网站,囊括先进的数据银行、出版集成商和图书馆目录的访问站点与全文传递集成系统。在 GetInfo 背后是 TIB 所提供的一整套科学技术专业信息服务。

作为欧洲大型图书馆入口之一的 GetInfo 为人们提供对科研数据进行访问获取的服务。这种新型内容的服务包括两种途径:

- 在 GetInfo 和联合目录(Gemeinsamen Bibliotheksverbund,GBV)中进行检索。用户寻找特定主题的信息时,通常会对某些科学家所收集的数据集感兴趣。他们可以直接使用永久标识符(这里指的是 DOI 数字资源唯一标识符),在图书馆目录上直接看到作为独立对象的科研数据。

在 GetInfo 所显示的搜索结果之中,科研数据位居中央。经过解析永久识别符,人们能在页面上从元数据看到数据中心,从数据中心获得数据集,这就像下载某条记录或者整个数据的链接一样。这套工作流程遵守 DOI 在科学期刊上的利用方式与功能效果。

- 科研数据的可视化。在 GetInfo 上提供类似地质科研数据的可视地图。用户在 GetInfo 解开 DOI 域

名后,可以进入如 PANGAEA(盘古大地)这样的地质数据中心的网页,进入数据中心开始访问数据。

3　科学数据的永久标识:DOI 服务

整合出版物的数据是科学合作的一个重要组成部分,它允许科研成果接受检验而且还能积极地交流科研知识。

尽管科学领域普遍认为数据应为跨学科所利用,然而,科学社群认识到目前很少有人对数据预处理、相关文献整理、确保数据质量上进行额外工作。

发表科学成果的传统形式是在专业期刊上发布,但通常没有基础数据的发布。这些传统出版进入"引用索引"后,使得科学家可以受到绩效评价。然而,这套索引到目前为止并不包括发布数据在内。

德国科学基金会(Deutsche Forschungsgemeinschaft,DFG)委托 TIB 所实行的《科研数据的发布与引用》项目计划,已经成功地测试了以 DOI 域名和 URNs 注册的基础设施来处理科学数据[4]。在 TIB 的系统中已经可以看到取自地理科学的超过 650 000 条记录的永久识别符。

作为识别符的 DOI 的应用,可以链接一篇科学论文和论文所分析的科研数据。文章和数据集通过他们各自的 DOI 域名,能够共同被引用。这种引用和链接的形式所提供的科研数据成为直接指向科学论文的附加数据(supplementary data)引用模式。其范例如下:

- 附加数据的引用标识:

Kuhlmann, H. et al. (2009): Age models, iron intensity, magnetic susceptibility records and dry bulk density of sediment cores from around the Canary Islands. PAMGEA, Bremen

doi:10.1594/PANGAEA.727522

- 关联到相关的科研论文:

Kuhlmann, Holger; Freudenthal, Tim; Helmke, Peer; Meggers, Helge (2004): Reconstruction of paleoceanography off NW Africa during the last 40,000 years: Influence of local and regional factors on sediment accumulation. In: Marine Geology, 207 (1 – 4), 209 – 224

doi:10.1016/j.margeo.2004.03.017

4　科学数据的全球引用:DataCite

基于 TIB 作为 DOI 注册代理者处理科研数据的经验,众多科技图书馆和信息中心在 2009 年 12 月共同成立 DataCite 国际联盟。它的目标是:让科学家在互联网上访问科研数据,接受科研数据作为独立而且受到引用的科研对象,从而遵循《优良的科学实践原则》[5]。已有来自 10 个国家的 15 个伙伴共同搭起 DataCite 的屋顶,它们是:大英图书馆、法国科技信息中心(Institute de I' Information Scientifique et Technique,INIST)、丹麦科技信息中心、荷兰的 Delft 工业大学图书馆、加拿大科学技术信息中心(Canada Institute for Scientific and Technical Information, CISTI)、美国能源部的科技信息办公室(Office of scientific and Technical Information in US Department of Energy, OSTI)、美国加州数字图书馆、美国普渡大学、澳大利亚国家数据服务中心(Australian National Data Service, ANDS)、瑞典国家数据服务中心(Swedish National Data Service, SND)以及苏黎世联邦技术学院。德国合作伙伴来自 TIB 的 Goportis 伙伴们:德国医学中央图书馆(Deutsche Zentralbibliothek für Medizin, ZB MED)、德国经济科学图书馆以及莱布尼兹社会科学院 GESIS(Leibniz - Institut für Sozialwissenschaften GESIS)。

DataCite 自成立以来就将自己作为数据中心、科学家、出版商和图书馆之间的中介。它在以下三个领域中提供服务:

- DataCite 是数字资源唯一标识符(DOI)的官方注册机构以及国际 DOI 基金会(IDF)的成员。截至 2011 年 1 月 DataCite 已为超过 100 万件数字对象进行了 DOI 名称注册。

- DataCite 积极发展标准规范与最佳实践的交流以及制定工作流程,所以在 DataCite 合作伙伴中有一个工作小组,负责与这个社群交流一套共通元数据架构,以囊括所有的科研数据引用信息[6]。DataCite 持续积极参与新成立的 CODATA 数据引用标准和实践的指导组(www.codata.org/taskgroups/TGdatacitation/index.html)的工作,制定有关科研数据引用的建议。不断与出版商讨论合作的可能性:用建立数据引用索引来鼓励研究人员,使他们的数据可用。

- DataCite 注册的所有数据,自 2011 年起,都将集中在一个可供访问的中央元数据库。这个数据库的内容可免费使用,并且可通过第三方,如汤森路透的 Web of Science 等站点访问获取。

所有这些活动都遵循 DataCite 的创始原则:科学是全球性的,因此它需要一个全球范围的伙伴;愿意与其他全球化机构,如出版商等,共同发展合适的标准规范。成员伙伴之中许多科学家们热衷于本地推广活

动,在他们的若干科研领域、图书馆和资助机构内建立网络。DataCite 得益于许多当地科学家的服务举措,而受到当地合作机构伙伴们的支持。

5 面向未来的图书情报工作服务

在图书馆目录中所显示的各种内容,将对图书馆和图书馆员产生新的意义[7]。高校与专业图书馆可以把自己作为服务设施,提升大学或者科研院所的科研数据访问能力;或者,与区域研究中心合作建立数据中心,在图书馆目录中导入资源链接。

图书馆学对存储和维护科研数据集而言也很重要。在英语国家有《数据馆员》这类著作,德国的波茨坦专科学校已经首次开设这门课程[7]。图书馆的工作人员可以积极参加这类机构为知识对象和出版领域所进行的培训,以便在必要之时能适当处理科研数据。

参考文献:

[1] National Science Foundation. Cyberinfrastructure vision for 21st century discovery [OL]. [2013-03-08]. http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf.

[2] Hey T, Tansley S, Tolle K. The fourth paradigm: Data-intensive scientific discovery [M]. Redmond, WA: Microsoft Research, 2009: ixx.
【译者注:该书已有中文版:潘教峰,张晓林,等译. 第四范式:数据密集型科研[M]北京:科学出版社, 2012.】

[3] Inger S, Gardner T. How readers navigate to scholarly content [OL]. [2013-03-08]. http://www.sic.ox14.com/howreadersnavigatetoscholarlycontent.pdf

[4] Heery R, Lyon L. Research and advanced technology for digital libraries [M]//Brase J. Using Digital Library Techniques – Registration of Scientific Primary Data. Berlin: Springer, 2004:488-494.

[5] Brase J. DataCite – A global registration agency for research data [OL]. [2013-03-08]. http://ieeexplore.ieee.org/stamp/stamp.jsp? tp=&arnumber=5361881.

[6] DataCite. DataCite metadata scheme for the publication and citation of research data [OL]. [2013-03-08]. http://schema.datacite.org/.

[7] Pampel H, Hobohm H C, Bertelmann R. Data librarianship-Rollen, Aufgaben, Kompetenzen [OL]. [2013-03-08]. http://hdl.handle.net/10760/14896.

Development Thinking and Practice on the German National Science and Technology Library

Jan Brase[1] Uwe Rosemann[1] Irina Sens[1] (Authors) Ku Liping[2] (Translator)

[1]German National Library of Science and Technology, Hannover 20167

[2]National Science Library, Chinese Academy of Sciences, Beijing 100190

[Abstract] This paper briefly introduces the new development of research data application and the new challenge of research data rights management. It describes a scalable framework of the German National Science and Technology Library (TIB) and its practical achievements on the research data portal GetInfo, the permanent identification service DOI and the global data citation service DataCite, and then discusses the need of data librarian coaching. Those services can being to enhance the non-textual resource retrieval, to identify digital data resource, to make research data be official cited and to improve library value.

[Keywords] data citation digital object identifier research data integrated portal rights and management on research data metadata scheme